IEEE *Access*

Multidisciplinary : Rapid Review : Open Access Journal

RESEARCH ARTICLE

A Method for Detecting Small Slow Targets on Sea Surface Based on FC-MCNN

QING SUN[ID], JING ZHAO, CHUNLING XUE[ID], AND XUETING YANG

School of Mathematics and Information Science, Baoji University of Arts and Sciences, Baoji 721000, China

Corresponding author: Qing Sun (sunq2009@126.com)

ABSTRACT Detecting small slow targets on sea surface has always been a hot topic in radar target detection. Classical target detection methods based on single input mode are particularly prone to losing the phase information of the radar echo signal, which affects the detection performance of the radar. To solve the problem of phase information loss, a False alarm Control method of Multi-input Convolutional Neural Network (FC-MCNN) is proposed in this article. First, several classic Convolutional Neural Network (CNN) models are compared and analyzed. Then, a Multi-input Convolutional Neural Network (MCNN) is designed to extract more useful information from radar echo signal. Simultaneously, in order to improve the generalization performance of the model and control the false alarm probability, a controllable false alarm Support Vector Machine (SVM) is used to modify the final output module. In addition, the datasets are also preprocessed before being input into the proposed model. Finally, the performance of the proposed method is verified by the IPIX measured data sets. The results show that the proposed method has a higher detection probability and better computational performance than the traditional CNN detection methods.

INDEX TERMS Target detection, constant false alarm rate (CFAR), signal processing, convolutional neural network (CNN), support vector machine (SVM).

I. INTRODUCTION

When marine radar detects sea surface targets, the radar receiver not only obtains partial scattered echoes of the target, but also contains a large number of scattered echoes from the illuminated sea surface, namely sea clutter [1]. Sea clutter is a type of interference clutter relative to targets, which seriously affects the detection performance of sea surface targets [2], [3]. Especially for detecting small slow targets on sea surface, such as floating ice, navigational buoys, small boats, frogmen, aircraft wreckage etc., they usually have small radar scattering cross-sections (RCS) and weak scattering echoes [4], [5]. If the environment is affected under the complex sea states, the target will be more difficult to detect effectively. Therefore, radar detection of small slow

targets under the complex sea states is a topic worthy of in-depth research.

Sea surface target detection methods can be classified into two categories based on their adaptive ability: artificial target detection methods and intelligent target detection methods. Artificial target detection methods have low adaptability and are difficult to apply to new environments when the target characteristics change [6]. With the continuous development of deep learning, the target detection method based on intelligent feature extraction not only shows better feature learning ability compared with the traditional target detection methods based on artificial feature selection, but also can get rid of the limitations of artificial feature selection and better self-adapt to learn high-dimensional features [7], [8]. For instance, target detection in machine vision, SAR image and other fields by combining the knowledge of related fields [9], [10], [11], [12], [13], [14], [15]. Although deep learning can improve the accuracy of image recognition

The associate editor coordinating the review of this manuscript and approving it for publication was Gerardo Di Martino[ID].

to a certain extent, excessively deep convolutional layers require a large amount of computing resources as the computational complexity increases, which requires high device performance and is difficult to promote and apply [16]. In view of the problems faced by deep learning, some scholars have carried out research on lightweight convolutional neural networks (CNNs) by optimizing the network structure and meeting certain detection requirements [17], [18], [19], [20].

On the basis of maintaining a certain detection performance, the lightweight CNNs can reduce the complexity of the model through pruning and other optimization design, so that the computational performance of the training model can be significantly improved [17], [18], [19]. Therefore, the application of lightweight CNNs has potential application value. Based on three typical lightweight CNNs, Simonyan and Zisserman [21] realized the detection and classification of micro-moving targets on sea surface by transfer learning method, and took four types of moving targets as research objects to verify the advantages of the CNN-based method compared with traditional target detection methods. On the basis of redesigning the fully connected network structure, Sun et al. [6] proposed an anchor-free method for detecting ship targets in high-resolution SAR images. Zhou et al. [22] improved a YOLOv3 model based on the pyramid structure, and applied the improved model to the multi-target detection of radar images in sea clutter. This method overcomes the limitations of the two-stage detection method and improves the accuracy of multi-target detection. However, because the improved method increases the network depth, the model parameters during training reach 246 M, which is difficult to realize engineering application. In addition, several models, such as VggNet [23], [24], GoogleNet [25], ResNet [26], and AlexNet [17], [27], have also been widely applied in radar image processing, but the above methods are mainly based on single input of the radar image, and relatively little research has been done on multi-source input, ignoring the information contained in the radar echo signal in different domains.

In the sea clutter environment, the CNN-based target detection or classification method usually reconstructs the pseudo-color image dataset based on the Time-Frequency (TF) transformation of the radar echo signal, and then uses the dataset to train the CNN model to extract high-dimensional features for target detection or classification. However, when generating the pseudo-color image dataset, the CNN can only obtain the boundary features of the image, and the phase information of the complex matrix obtained by TF transformation is lost. In addition, the statistical characteristics of radar echo signals are not considered in pseudo-color images. Considering that the phase information contained in the TF transformation complex matrix is an effective source of information for target detection, the proposed method extracts the real and imaginary parts of the complex matrix based on the Short-Time Fourier Transform (STFT), and generates the corresponding pseudo-color image dataset as the input images. In the mean time, the statistics of radar echo signals are also considered as input vector. Finally,

the target detection is carried out by combining the false alarm controllable Support Vector Machine (SVM) classifier.

Aiming at the problem that small slow targets are difficult to detect, we propose a False alarm Control method of Multi-input Convolutional Neural Network (FC-MCNN). Our main work is summarized as follows:

- To overcome the problem of phase information loss caused by single input and insufficient image information extraction, a multi-input network structure is designed to increase the amount of input information and facilitate the mining of multi-domain features.
- To preserve useful information between different layers, a Skip convolutional layer is employed, which allows the network to automatically choose whether to extract features from multiple convolutional layers or skip them directly, effectively preserving information from multiple layers.
- To improve the generalization performance of the proposed model, a SVM module is adopted, based on the high-dimensional features extracted by Multi-input Convolutional Neural Network (MCNN), and useful information is reasonably selected, effectively improving the generalization and computational performance of the model. The total number of model parameters is only 2.6 M.
- To control the false alarm probability output by the neural network, a Constant False Alarm Rate (CFAR) module is designed to achieve target detection function at a given false alarm probability.

The rest of this paper is organized as follows. In Section II, several types of classical network structure models are introduced. Then, based on the analysis of the typical network structure model, the target detection method with controllable false alarm and multiple input are designed in Section III. For the needs of input and better detection of small slow targets, Section IV further introduces the processing of input data. Section V uses the measured datasets to verify the effectiveness of the proposed method. Finally, conclusions are drawn in Section VI.

II. CONVOLUTIONAL NEURAL NETWORK MODEL

The traditional radar target detection method based on convolutional neural network (CNN) is to obtain classification results by mapping the input data, and the formula can be expressed as follows:

$$\mathbf{C}_{pred} = f_{softmax}(f_{fc}(f_{conv}(\mathbf{X}))) \tag{1}$$

where $\mathbf{C}_{pred}$ represents classification result, $f_{softmax}$ represents classification operation, f_{fc} represents full link operation, f_{conv} represents convolution operation, and $\mathbf{X}$ represents input data.

At present, the widely used lightweight CNNs include LeNet, AlexNet and GoogleNet, as well as a series of lightweight networks developed on their basis [17], [25], [27], [28], [29], [30]. The following takes three typical lightweight
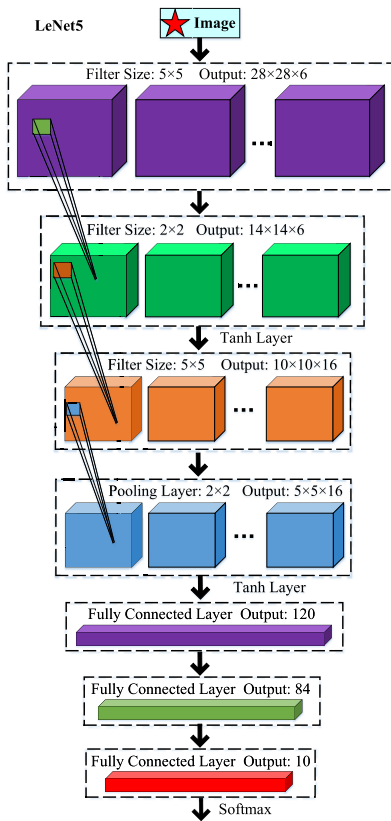
FIGURE 1. Diagram of the LeNet5 basic framework.

FIGURE 2. Diagram of the AlexNet's basic framework.

where i is the output class, j is the fully connected node, y_i is the error Radial Basis Function (RBF) value of class i, and w_{ij} is the parameter vector. The final output is a Euclidean radial basis unit, one unit per class, and each output Euclidean radial basis unit calculates the Euclidean distance between the input vector and the parameter vector, the smaller the value, the more likely the result is in the i class. Thus, the output of an error RBF can be regarded as a penalty measure for matching a model to the associated class of the error RBF.

CNNs as examples to briefly introduce the basic architectures of the CNNs.

A. LeNet NETWORK

LeNet is the earliest CNN and has been successfully applied to the recognition of handwritten digits. Its architecture is characterized by the distribution of image features on the entire image. By using convolution kernel with learnable parameters, similar features are extracted at multiple locations for the recognition of handwritten digits [28]. Represented by LeNet5 in the LeNet series, the model mainly contains three types of parameter layers, namely two 5×5 convolution kernel layers, two fully connected layers and a Softmax layer, with each convolution layer followed by a pooling layer. The basic structure is shown in Fig. 1.

For the final fully connected layer, LeNet5 is output by the Tanh function, which is formulated:

$$\tanh x = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{2}$$

where x is the sum of the dot product of the output vector and the weight vector of the fully connected layer.

A state is generated by the Tanh function, and then a Gaussian connection is formed, and the connection is calculated as follows:
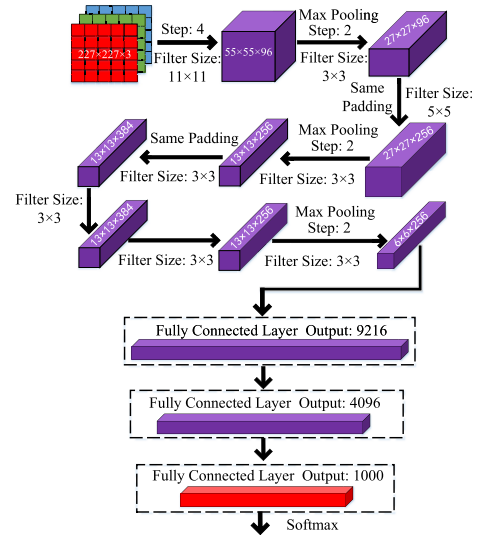
$$y_i = \sum_j (x_j - w_{ij})^2 \tag{3}$$

B. AlexNet NETWORK

The AlexNet has promoted the widespread application of the CNNs, introducing ReLU activation function, overlap pooling, local response normalization, data enhancement, and random discard operations, which not only increases training speed and reduce overfitting, but also improves computational accuracy. In addition, the network utilizes the Compute Unified Device Architecture (CUDA) to accelerate the training process of the CNNs in parallel, which fully exploits the powerful computing performance of Graphics Processing Unit (GPU). Its basic framework is shown in Fig. 2 [17]. The calculation formula of ReLU activation function in the network is as follows:

$$f(x) = \max(0, x) \tag{4}$$

C. GoogleNet NETWORK

Based on optimizing specific fully connected layers, GoogleNet designed the Inception module. The basic structure of Inception module is shown in Fig. 3, where e_1, e_2, e_3, e_4, e_5 and e_6 are hyperparameters [25]. Three types of convolution kernel filters are adopted in this module. On the basis of 1×1 convolution kernel dimensionality reduction sampling, a parallel path is constructed, and the corresponding filter convolutional layer is obtained by combining 3×3 convolution kernel, 5×5 convolution
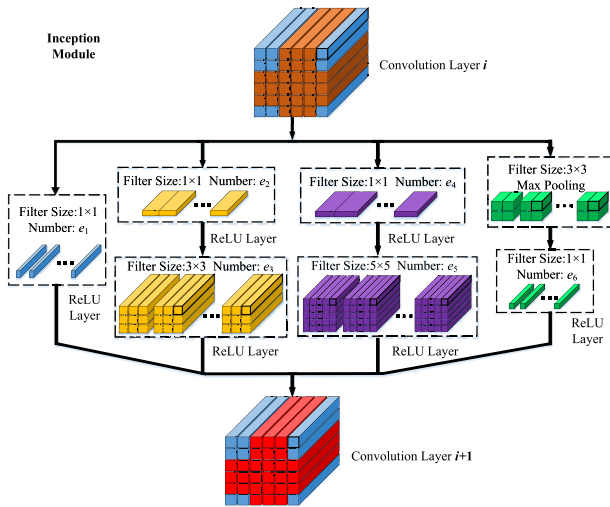
FIGURE 3. Diagram of the inception module structure.

FIGURE 4. Diagram of the MCNN basic framework.

kernel, ReLU activation function and maximum pooling. Finally, each convolutional layer is deeply connected in the channel dimension, and its core design concept is to separate the spatial correlation explicitly, so as to achieve the decoupling operation of channel correlation and spatial correlation.

III. TARGET DETECTION METHOD BASED ON FC-MCNN

The essence of small slow target detection on sea clutter is to decide whether there is a target in the radar echo signal. The classical target detection is a Constant False Alarm Rate (CFAR) detector. However, under complex sea states, there is a problem of insufficient deep information mining for radar echo signals. Especially when using traditional methods to detect small slow targets, it is easy to misidentify strong sea clutter as the target, resulting in a high false alarm rate, or misidentify the radar echo signal of weak targets as sea clutter, leading to omission of real targets. The traditional CNN method extracts high-dimensional features from single input data, and then determines whether there are small slow targets in the radar echo signal. However, other characteristics included in the radar echo signal have not been fully utilized. Therefore, based on the preprocessing of the radar echo signal, this paper processes the radar echo signal from different perspectives, and then inputs them as multi-domain datasets. The specific network model framework is as follows.

A. DESIGN OF MULTI-INPUT CONVOLUTION NEURAL NETWORK

Classical CNNs are mostly based on extracting features from a single image input for target detection, while there is less research on target detection under multi-input conditions, so there are few network structures that can be used. In order to mine the information contained in different domains of radar echo signals, a Multi-input Convolutional Neural
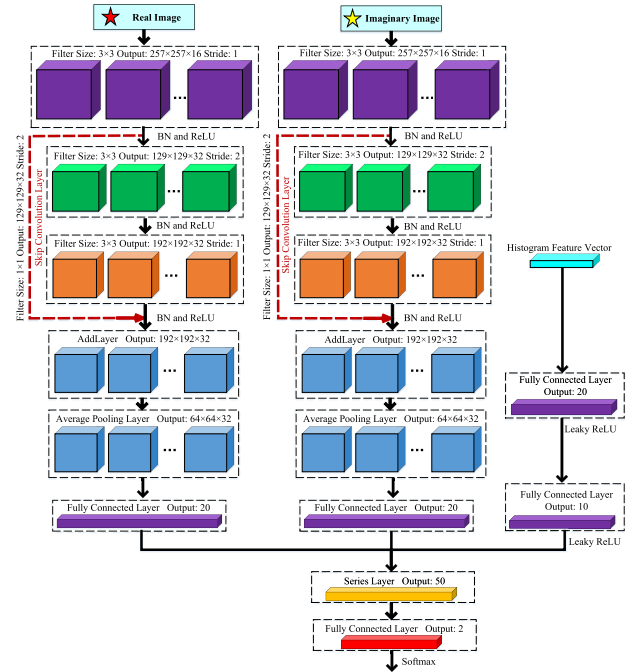
Network (MCNN) framework is designed to meet the needs of marine target detection under multi-source information input conditions. The network mainly includes three parts: input layer module, hidden layer feature extraction module and result output module.

Based on the designed parallel framework, the high-dimensional features contained in the input data are extracted respectively, and then the network results are fused before output. Subsequently, the trained MCNN model is obtained. The framework is shown in Fig. 4, and the specific module design is as follows.

1) INPUT LAYER MODULE DESIGN

The input layer module consists of three parts: real part input layer, imaginary part input layer, and histogram feature vector input layer. Different from the traditional input layer, when designing the input layer, this paper considers that the data obtained by STFT is a complex matrix, and the real and imaginary parts of the complex matrix are orthogonal components, which are independent of each other and contain different echo information. If the single input layer method is adopted, only the TF power spectrum information is used, and the multidimensional information in the complex matrix will be lost. Therefore, on the basis of analyzing the STFT characteristics, this paper obtains two sets of matrices by separating the real and imaginary parts, and then enters them into MCNN in parallel to extract the different information contained in the real and imaginary parts. In addition, considering the statistical characteristics of the radar echo signal, this information is further extracted and input into the MCNN to improve the performance of target detection.

2) HIDDEN LAYER FEATURE EXTRACTION MODULE DESIGN

In the hidden layer feature extraction module, a parallel multi-layer network framework is designed, as shown in Fig. 4. The hidden layer feature extraction module mainly includes convolution layer, Skip convolution layer, ReLU layer, pooling layer and fully connected layer. To extract high-dimensional features of input vector generated by histogram, a network framework consisting of full connection layer and leaky ReLU layer is adopted.

3) RESULT OUTPUT MODULE DESIGN

On the basis of extracting multi-domain high-dimensional features separately, the features are fused, and then the fused features are reduced in dimensionality using a full connection layer to obtain the reduced deep features, which are then input into Softmax. Softmax calculates the probability table of the output class, and predicts the classification result based on the probability table. Secondly, cross-entropy is used to calculate the loss between the predicted value and the true value, and then the model parameters are updated according to the Stochastic Gradient Descent with Momentum (SGDM) method until the given allowable error is satisfied or the maximum iteration steps is reached. Finally, the trained MCNN model is used to extract high-dimensional features. The main formulas used in this module include Softmax layer output formula, cross-entropy loss formula and SGDM formula [31].

The output formula in the Softmax layer is:

$$Softmax(y_i) = \frac{e^{y_i}}{\sum_{j=1}^{N} e^{y_j}} \tag{5}$$

where $Softmax(y_i)$ represents the output probability of the y_i class and N represents the number of classes. Softmax converts the output of this module into a probability distribution, which in turn provides data for the next step of calculating the distance between the true and predicted values using cross-entropy.

The calculation formula for cross-entropy loss is:

$$H(p, q) = -\sum_{x} p(x) \log q(x) \tag{6}$$

where x represents the input class, p represents the probability of the true value, and q represents the probability of the predicted value. The smaller the cross entropy is, the closer the distribution of the two probabilities is.

The formula of SGDM is:

$$\theta^{l+1} = \theta^l - \alpha \nabla E(\theta^l) - \beta(\theta^l - \theta^{l-1}) \tag{7}$$

where θ represents the input model parameter, l represents the iterative step, α represents the learning rate, E represents the loss function, ∇ represents the gradient, and β represents the influence coefficient of the previous parameter and the current parameter on the next parameter.
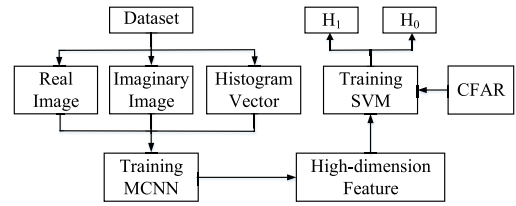

FIGURE 5. Target detection method based on FC-MCNN.

B. SVM MODULE DESIGN WITH CFAR

Although the MCNN model can use the Softmax layer to output the predicted classification results, its false alarm rate is usually calculated after the classification, and its false alarm rate is an output quantity, which cannot guarantee the use of MCNN to detect the sea surface target under the given CFAR. In addition, in the Softmax output module, all the data input from the front layer is required to participate in MCNN decision-making, so that each input data from the front layer will affect the result of MCNN prediction and classification. Considering that the objective function of SVM is convex, the optimization goal is to minimize structural risk rather than to minimize empirical risk, and under the set false alarm probability, SVM only selects a few features as support vector to determine the hyperplane, which has better model generalization ability and classification accuracy. Therefore, on the basis of the designed MCNN model, this section takes the high-dimensional features extracted from the MCNN model as the inputs of SVM, combined with the set false alarm probability, and further trains the SVM classifier to obtain a radar target detection method based on FC-MCNN. The detection process of the proposed method is shown in Fig. 5.

IV. PREPROCESSING METHOD OF RADAR ECHO SIGNAL

The FC-MCNN designed in this paper adopts two image input layers and one feature vector input layer on the basis of considering TF domain characteristics and amplitude macro-statistical characteristics. Therefore, to obtain better high-dimensional features, it is necessary to perform STFT and normalized histogram processing on the data before inputting radar echo signals into the model.

A. SHORT-TIME FOURIER TRANSFORM

STFT can not only process stationary signals, but also non-stationary signals. Its basic idea is to assume that the radar echo signal under a certain window function is a stationary signal. By moving the window function, the spectrum of the local radar echo signal at different times is calculated, and then the TF spectrum is obtained. It consists of mutually orthogonal real and imaginary parts. STFT is defined as [32]:

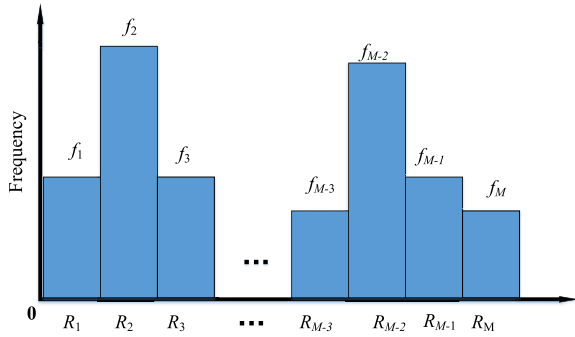$$\text{STFT}(t, f) = \sum_{\tau=-\infty}^{\infty} S(\tau)w(t-\tau)e^{-j2\pi f\tau} \tag{8}$$

FIGURE 6. Diagram of normalized histogram.

FIGURE 7. ASCR of the selected datasets.

where $S(\tau)$ is the input signal. $w(\tau)$ is the window function, which is a time reversal and has an offset of t moments. $STFT(t,f)$ is a two-dimensional function of time and frequency, according to which the TF analysis of radar echo signal can be carried out. It should be noted that once the window function is determined, the resolution of the STFT is also determined. In this paper, Hamming window is selected as the window function, and other parameters of window function are determined by the length of input signal.

B. NORMALIZED HISTOGRAM

Histogram is used to represent the relationship between frequency of occurrence of values in data and group distance. Taking the group distance as the horizontal coordinate and the frequency as the vertical coordinate, the histogram is obtained by calculating the relationship between the group distance and the frequency, which is an important feature of the data and reflects the macro-statistical characteristics of the data. The histogram includes the number of features to be counted, the number of group distances for each feature space, and the value range for each feature space.

For a group of radar echo signals, assuming that the echo data set is S, the group distance is R_m, $m = 1, 2, \cdots, M$, the value range of the feature space is $[S_{min}, S_{max}]$, and the normalized frequency value is $f(m)$, the definition formula of the normalized histogram can be obtained as follows:

$$f(m) = \frac{h(R_m)}{\sum_j h(R_j)} \quad (9)$$

where $h(R_m)$ represents the number of S corresponding to the group distance R_m, and $\sum_j h(R_j)$ represents the total number of S. The normalized histogram is shown in Fig. 6.

It is worth noting that for radar echo signals, a group of radar echo signals corresponds to a unique histogram when the group distance is fixed, but a histogram can correspond to multiple groups of radar echo signals.

V. EXPERIMENTAL RESULTS AND ANALYSIS

The datasets used in the experiment were the X-band IPIX datasets released by McMaster University in Canada, which were measured in two batches. The first batch datasets was in
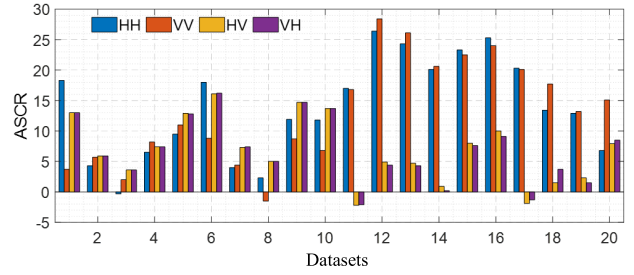
1993 for a floating ball with a radius of 1m as the detection target, and the second batch datasets was in 1998 for a slow-moving boat as the detection target [33]. According to the characteristics of the two batches of datasets, 10 groups of typical sea state and Average Signal-to-Clutter Ratio (ASCR) datasets are selected for each batch, for a total of 20 datasets, and the relevant parameters are detailed in Table 1 and Fig. 7. The datasets in Table 1 cover a variety of sea states, polarization modes and ASCR, and Fig. 7 shows the ASCR under various polarization modes in Table 1. The ASCR calculation formula is as follows [34]:

$$ASCR = 10 \log_{10} \frac{P_{\mathrm{T}} - P_{\mathrm{C}}}{P_{\mathrm{C}}} \quad (10)$$

where P_{T} represents the average power of the target bin and P_{C} represents the average power of the sea clutter bin.

TABLE 1. Description of the IPIX datasets.

No.	File Name	Target Rangebin	Guard Rangebin	Wave Height (m)
1	19931107_135603	9	8, 10, 11	2.1
2	19931108_220902	7	6, 8	1.1
3	19931109_191449	7	6, 8	0.9
4	19931109_202217	7	6, 8, 9	0.9
5	19931110_001635	7	5, 6, 8	1.0
6	19931111_163625	8	7, 9, 10	0.7
7	19931118_023604	8	7, 9, 10	1.6
8	19931118_162155	7	6, 8, 9	0.9
9	19931118_162658	7	6, 8, 9	0.9
10	19931118_174259	7	6, 8, 9	0.9
11	19980204_163113	24	23, 25, 26	-
12	19980204_202225	24	23, 25, 26	-
13	19980204_202525	7	6, 8, 9	-
14	19980205_171437	7	6, 8, 9	-
15	19980205_180558	7	6, 8, 9	-
16	19980212_195704	7	6, 8, 9	-
17	19980223_164055	31	30, 32, 33	-
18	19980223_173317	32	31, 33, 34	-
19	19980223_173950	29	28, 30-34	-
20	19980304_184537	21	20, 22	-

Other parameters are selected as follows: in the HH polarization mode dataset, 5,000 samples are generated from the sea clutter bin and the target bin with a window width of 512, and the total number of positive and negative samples
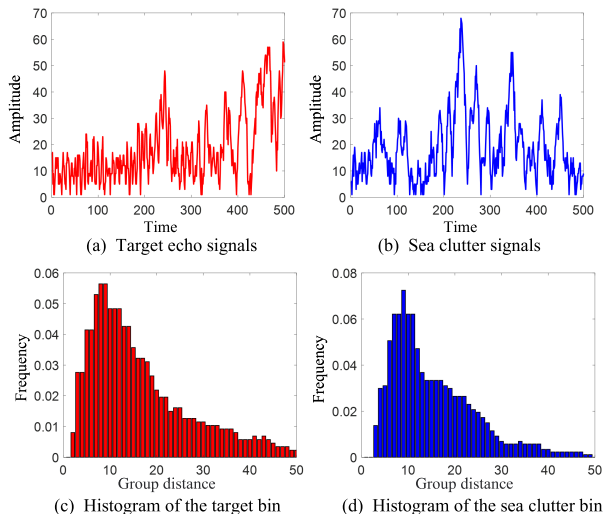
FIGURE 8. Radar echo signals and their histograms.

is 10,000, and then the samples are preprocessed, and finally the preprocessed samples are grouped into training set and test set according to the ratio of 8:2.

A. INPUT DATASET CHARACTERIZATION

The dataset 19931118_023604 is taken as an example to demonstrate the processing effect of the input dataset. Firstly, randomly intercept two samples containing the target bin and the pure sea clutter bin from the dataset 19931118_023604, as shown in Fig. 8. Among them, Fig. 8(a) and (b) are the original radar echo signals of the target bin and the sea clutter bin respectively, and Fig. 8(c) and (d) are the normalized histograms extracted from the target bin and the sea clutter bin respectively.

It can be seen from Fig. 8(a) and (b) that the amplitude of the target position and the pure sea clutter position fluctuates within a certain range, and the difference in amplitude is not obvious. However, after processing by normalized histogram, the two samples show different amplitude statistical distribution characteristics. The cosine similarity is further used to calculate the similarity of the two, and its value is 9.2×10^{-4}, approaching 0, that is, the normalized histograms of the target bin and the pure sea clutter bin are independent of each other, indicating that the distribution of the two has information that can be mined. The formula for calculating cosine similarity is as follows [35]:

$$S_{cos} = \frac{\sum_i A_i \times B_i}{\sqrt{\sum_i A_i^2} \times \sqrt{\sum_i B_i^2}} \qquad (11)$$

where A_i represents the probability value at the i-th group distance of the target bin histogram, B_i represents the probability value at the i-th distance of the pure sea clutter bin histogram, S_{cos} represents the cosine similarity between the target bin and the pure sea clutter bin, and the value range of cosine similarity is $[-1,1]$. The closer the cosine similarity
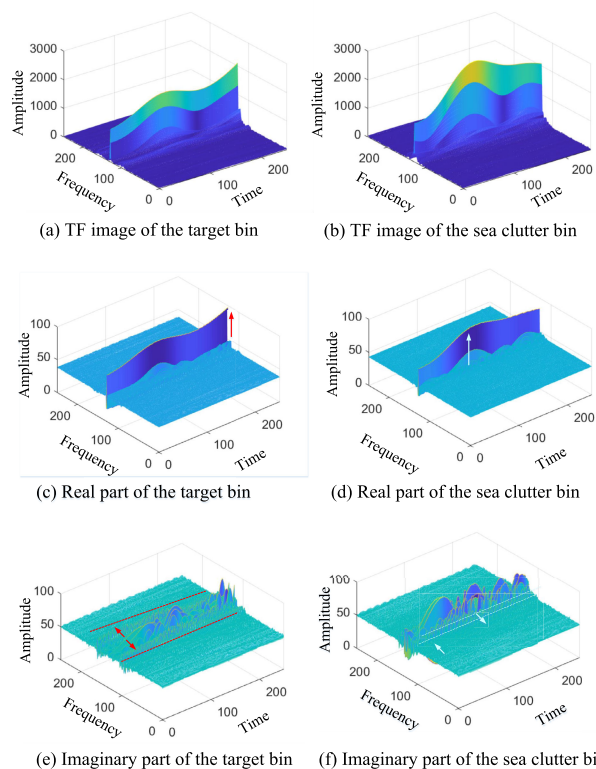

FIGURE 9. TF images of the radar echo signal.

value is to -1, the stronger the negative correlation is. The closer to 1, the stronger the positive correlation, and the closer to 0, the stronger the independence.

Secondly, STFT is carried out on the extracted radar echo signal to obtain the TF pseudo-color image and the corresponding pseudo-color images of the real and imaginary parts, as shown in Fig. 9. Among them, Fig. 9(a) is the TF pseudo-color image of the target bin, Fig. 9(c) and (e) are the pseudo-color images of the real and imaginary parts of the target bin, Fig. 9(b) is the TF pseudo-color image of the pure sea clutter bin, and Fig. 9(d) and (f) are the pseudo-color images of the real and imaginary parts of the pure sea clutter bin, respectively.

As can be seen from Fig. 9(a) and (b), the TF pseudo-color images after STFT mainly show the amplitude information, while the time and frequency plane characteristics are not significantly displayed. After separating the real and the imaginary parts of the TF pseudo-color image, it is obvious that the real part determines the longitudinal amplitude distribution characteristics of the pseudo-color image, as shown in Fig. 9(c) and (d). The imaginary part of the pseudo-color image determines the distribution horizontal amplitude characteristics of the pseudo-color image, as shown in Fig. 9(e) and (f). By comparing the pseudo-color images before and after the real and imaginary parts, it can be seen that the real and imaginary parts depict the characteristics of different dimensions and have potential mining value.

B. LOCAL FEATURE EXTRACTION AND ANALYSIS

In order to further investigate the effect of MCNN hidden layer feature extraction, the shallow features of the input image were extracted using the trained MCNN model based on the selected dataset in Section V, and the results are shown in Fig. 10. Among them, Fig. 10(a) and (e) are the features and strongest feature of the 16 channels of the first convolution layer extracted from the real part of the target bin, and Fig. 10(c) and (g) are the features and strongest feature of the 16 channels of the first convolution layer extracted from the imaginary part of the target bin. Fig. 10(b) and (f) are the features and strongest feature of the 16 channels of the first convolution layer extracted from the real part of the pure sea clutter bin, and Fig. 10(d) and (h) are the features and strongest feature of the 16 channels of the first convolution layer extracted from the imaginary part of the pure sea clutter bin.

From the comparison of the strongest feature of the real and imaginary parts of the target bin and the pure sea clutter bin, it can be seen that both the target bin and the pure sea clutter bin form bright double lines in the middle, indicating that the amplitude of the two is significantly strengthened in the central part, and some short ripples of different lengths are formed around the brightest lines. Moreover, the target bin is more concentrated in a certain deviation area on both sides, showing a linear distribution. The pure sea clutter bin distributes short ripples in a certain area off-center on both sides, and presents a zonal distribution. By comparing the first layer of convolution features extracted from the real and imaginary parts, it can be seen that the brightness of the feature images obtained in the imaginary part is stronger than that of the real part, which verifies that the real and imaginary parts contain different feature information, which is conducive to target detection.

C. DETECTION PERFORMANCE

To verify the detection performance of the proposed method, 20 groups of measured datasets in Table 1 are selected as experimental verification sets with a false alarm probability of 10^{-3}. In addition, the target detection performance of some classical CNN methods, such as LeNet5 [28], AlexNet [17], GoogleNet [25], MobileNetv2 [36], EfficientNetb0 [37] and Inceptionv3 [38], is further compared. The results are shown in Fig. 11, where the black solid line represents the detection probability of the proposed method. The average detection probability and model training time of each method are shown in Table 2. As can be seen from Fig. 11, the proposed method has a high detection probability in datasets ranging from 6 to 20, while the detection probability in datasets ranging from 1 to 5 is in the middle. The experiment results show that the detection performance of small slow targets on sea surface is not only affected by sea conditions such as sea state, but also by the geometry of the detection targets. Especially in the last 10 datasets, the detection probability of the proposed method is close to 100%. Obviously,
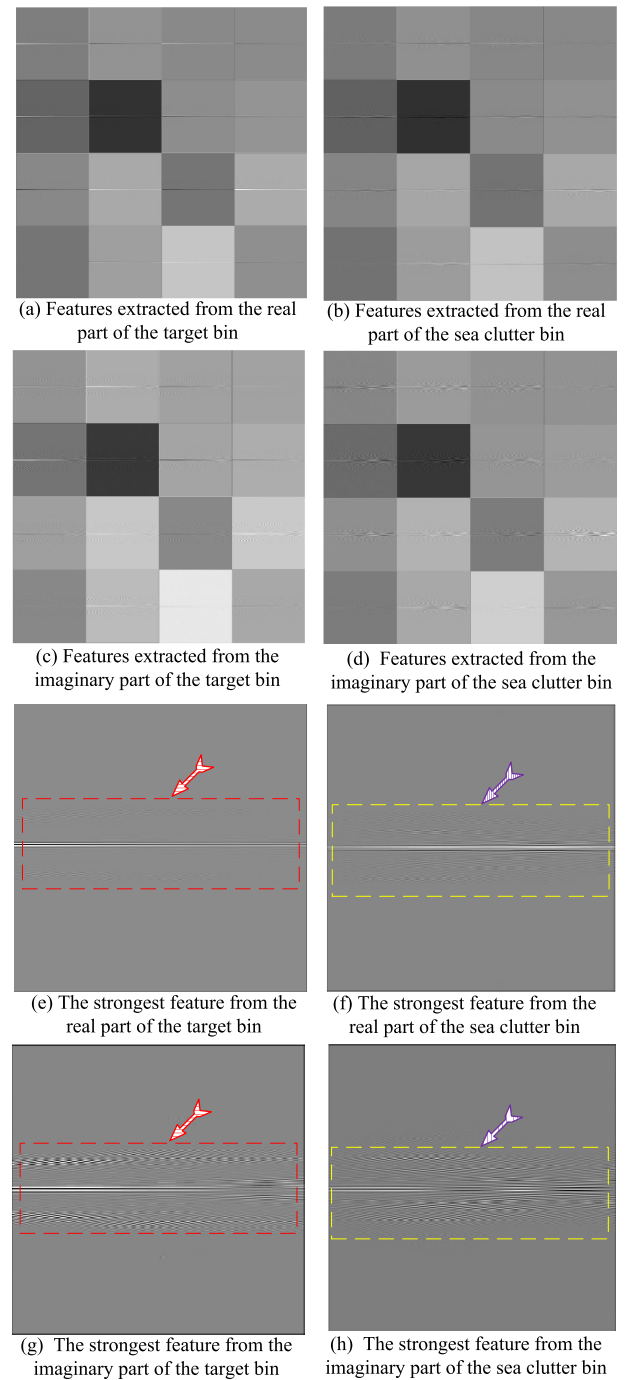

(a) Features extracted from the real part of the target bin

(b) Features extracted from the real part of the sea clutter bin

(c) Features extracted from the imaginary part of the target bin

(d) Features extracted from the imaginary part of the sea clutter bin

(e) The strongest feature from the real part of the target bin

(f) The strongest feature from the real part of the sea clutter bin

(g) The strongest feature from the imaginary part of the target bin

(h) The strongest feature from the imaginary part of the sea clutter bin

FIGURE 10. Hidden layer features of the input images.

due to the richer information from radar echo signal, the proposed method can mine more feature information for the second type of small slow target, resulting in better detection performance. In other words, this is because the proposed method makes full use of the input multi-source information, mines the high-dimensional features of the input data, and combines SVM for information fusion, so as to obtain better comprehensive detection performance under the set CFAR. As can be seen from Table 2, the proposed
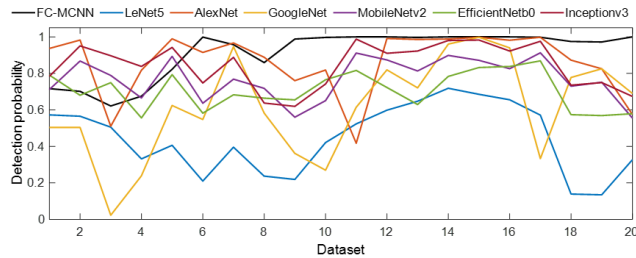
FIGURE 11. Detection probability of seven methods in the selected datasets.

TABLE 2. Detection performance of the seven methods.

Method	Detection Probability (%)	Training Time (min)	Model Size (M)
FC-MCNN	**91.31**	4	2.6
LeNet5	44.33	**3**	0.6
AlexNet	86.10	7	61
GoogleNet	61.41	16	7
MobileNetv2	77.03	42	3.5
EfficientNetb0	70.67	54	5.31
Inceptionv3	84.44	18	23.9

TABLE 3. The influence of different features on the detection performance.

No.	Input	Detection probability (%)	Sort
1	Real	85.80	5
2	Imaginary	78.78	6
3	Histogram	59.54	7
4	Real+imaginary	88.04	4
5	Real+histogram	89.58	2
6	Imaginary+histogram	89.08	3
7	Real+imaginary+histogram	**91.31**	**1**

method adopts a multi-input lightweight network design, the model size is only 2.6 M, and parallel processing is adopted, which significantly reduces the complexity of the model and improves the computational efficiency. The average training time of the proposed method is only 4 min, which is much less than other CNN detection methods, slightly higher than LeNet5, but has a higher detection probability.

To analyze the impact of different input features on the detection performance of the proposed method, this paper further conducted ablation experiments. The average detection probability of different inputs and their combinations was validated using 20 selected datasets, as shown in Table 3. From lines 1 to 3, it can be seen that the detection probability of the only real input is the highest, followed by the imaginary and histogram, indicating that the real part has a good ability to partition relative to the other two features. According to lines 4 to 6, in the case of dual feature inputs, the combination of dual inputs with histogram participation is most beneficial for performance improvement, followed by the combination of real and virtual. Compared to lines 2, 4, and 6, the imaginary part

also contains information different from the real part and histogram. When the real part, imaginary part and histogram are input simultaneously, the detection probability is 91.31%. The above experimental results indicate that the three feature inputs selected by the proposed method are effective, and that a better comprehensive detection performance is achieved by combining SVM and CFAR.

VI. CONCLUSION

Aiming at the problems of insufficient target feature information extracted by traditional CNNs, large training parameters, high computational complexity, and difficult to control the false alarm probability, a small slow target detection method based on FC-MCNN is proposed in this paper. Subsequently, comparative experiments were conducted to verify the detection performance of the proposed method under general sea conditions. However, the real environment in which small slow targets are located is usually more complex. In special scenarios, such as rainfall or snowfall, the radar echo signal will change, making it more difficult to detect small slow targets. Therefore, for target detection in special scenarios, the detection performance of the proposed method still needs further verification. In the future, detecting small slow targets in complex sea states has important research value and is also highly challenging.

REFERENCES

[1] C. Xue, F. Cao, Q. Sun, J. Qing, and X. Feng, "Sea-surface weak target detection based on multi-feature information fusion," *J. Syst. Eng. Electron.*, vol. 44, no. 11, pp. 3338–3345, 2022.

[2] X. Chen, W. Huang, C. Zhao, and Y. Tian, "Rain detection from X-band marine radar images: A support vector machine-based approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 2115–2123, Mar. 2020.

[3] X. Chen and W. Huang, "Identification of rain and low-backscatter regions in X-band marine radar images: An unsupervised approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 4225–4236, Jun. 2020.

[4] S. Xu, J. Zheng, J. Pu, and P. Shui, "Sea-surface floating small target detection based on polarization features," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 10, pp. 1505–1509, Oct. 2018.

[5] P. Meiyan, S. Jun, Y. Yuhao, L. Dasheng, and Y. Junpeng, "M-FCN based sea-surface weak target detection," *J. Syst. Eng. Electron.*, vol. 32, no. 5, pp. 1111–1118, Oct. 2021.

[6] Z. Sun, M. Dai. X. Leng. Y. Lei. B. Xiong, K. Ji, and G. Kuang, "An anchor-free detection method for ship targets in high-resolution SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7799–7816, 2021.

[7] N. Su, X. Chen, J. Guan, Y. Huang, and N. Liu, "One-dimensional sequence signal detection method for marine target based on deep learning," *J. Signal Process.*, vol. 36, no. 12, pp. 1987–1997, 2021.

[8] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, M. Hasan, B. C. Van Essen, A. A. S. Awwal, and V. K. Asari, "A state-of-the-art survey on deep learning theory and architectures," *Electronics*, vol. 8, no. 3, p. 292, Mar. 2019.

[9] G.-Q. Li, Z.-Y. Song, and Q. Fu, "A convolutional neural network based approach to sea clutter suppression for small boat detection," *Frontiers Inf. Technol. Electron. Eng.*, vol. 21, no. 10, pp. 1504–1520, Oct. 2020.

[10] Z. Baird, M. K. Mcdonald, S. Rajan, and S. J. Lee, "A CNN-LSTM network for augmenting target detection in real maritime wide area surveillance radar data," *IEEE Access*, vol. 8, pp. 179281–179294, 2020.

[11] S. Wang, Z. Chen, S. Du, and Z. Lin, "Learning deep sparse regularizers with applications to multi-view clustering and semi-supervised classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 5042–5055, Sep. 2022.

[12] Z. Zhao, X. Li, H. Liu, and C. Xu, "Improved target detection algorithm based on libra R-CNN," *IEEE Access*, vol. 8, pp. 114044–114056, 2020.

[13] Y. Yan and H.-Y. Xing, "A sea clutter detection method based on LSTM error frequency domain conversion," *Alexandria Eng. J.*, vol. 61, no. 1, pp. 883–891, Jan. 2022.

[14] Z. Sun, X. Leng, Y. Lei, B. Xiong, K. Ji, and G. Kuang, "BiFA-YOLO: A novel YOLO-based method for arbitrary-oriented ship detection in high-resolution SAR images," *Remote Sens.*, vol. 13, no. 21, p. 4209, Oct. 2021.

[15] Z. Lin, K. Ji, X. Leng, and G. Kuang, "Squeeze and excitation rank faster R-CNN for ship detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 5, pp. 751–755, May 2019.

[16] M. Kang, K. Ji, X. Leng, and Z. Lin, "Contextual region-based convolutional neural network with multilayer fusion for SAR ship detection," *Remote Sens.*, vol. 9, no. 8, p. 860, Aug. 2017.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[18] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," 2017, *arXiv:1707.01083*.

[19] S. Ningyuan, C. Xiaolong, G. Jian, M. Xiaoqian, and L. Ningbo, "Detection and classification of maritime target with micro-motion based on CNNs," *J. Radars*, vol. 7, no. 5, pp. 565–574, 2018.

[20] X. Zhang, H. Wang, C. Xu, Y. Lv, C. Fu, H. Xiao, and Y. He, "A lightweight feature optimizing network for ship detection in SAR image," *IEEE Access*, vol. 7, pp. 141662–141678, 2019.

[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[22] L. Zhou, S. Y. Wei, Z. M. Cui, J. Q. Fang, X. T. Yang, and L. Yang, "Multi-objective detection of complex background radar image based on deep learning," *Sys. Eng. Electron.*, vol. 41, no. 6, pp. 1258–1264, 2019.

[23] U. Muhammad, W. Wang, S. P. Chattha, and S. Ali, "Pre-trained VGGNet architecture for remote-sensing image scene classification," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 1622–1627, doi: 10.1109/ICPR.2018.8545591.

[24] J. He, S. Li, J. M. Shen, Y. Liu, and P. Jin, "Facial expression recognition based on VGGNet convolutional neural network," in *Proc. Chin. Autom. Congr. (CAC)*, Nov./Dec. 2018, pp. 4146–4151 doi: 10.1109/CAC.2018.8623238.

[25] Y. Ge, S. Jiang, Q. Xu, C. Jiang, and F. Ye, "Exploiting representations from pre-trained convolutional neural networks for high-resolution remote sensing image retrieval," *Multimedia Tools Appl.*, vol. 77, no. 13, pp. 17489–17515, Jul. 2018.

[26] S. Targ, D. Almeida, and K. Lyman, "Resnet in Resnet: Generalizing residual architectures," 2016, *arXiv:1603.08029*.

[27] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size," 2016, *arXiv:1602.07360*.

[28] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[29] K. T. Ahmed, S. Jaffar, M. G. Hussain, S. Fareed, A. Mehmood, and G. S. Choi, "Maximum response deep learning using Markov, retinal & primitive patch binding with GoogLeNet & VGG-19 for large image retrieval," *IEEE Access*, vol. 9, pp. 41934–41957, 2021.

[30] Z. Yu, Y. Dong, J. Cheng, M. Sun, and F. Su, "Research on face recognition classification based on improved GoogleNet," *Secur. Commun. Netw.*, vol. 2022, pp. 1–6, Jan. 2022.

[31] M. Xia, T. Li, L. Xu, L. Liu, and C. W. de Silva, "Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 101–110, Feb. 2018.

[32] D. Gabor, "Theory of communication," *IEE Proc. London*, vol. 93, no. 73, p. 58, 1946.

[33] McMaster University, Hamilton, ON, Canada. (Oct. 2012). *The IPIX Radar Database [OL]*. [Online]. Available: http://soma.mcmaster.ca/ipix.php

[34] P. R. Massopust, *Fractal Functions, Fractal Surfaces, Wavelets*. New York, NY, USA: Academic, 1994, pp. 235–304.

[35] C. Y. Liu, S. F. Wang, K. X. Wang, S. Chen, and Q. Sun, "Non-local means filter method based on cosine similarity index," *J. Changchun Univ. Sci. Tech.*, vol. 44, no. 2, pp. 18–26, 2021.

[36] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Improved sqrt-cosine similarity measurement," 2017, *arXiv:1704.0486*.

[37] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," 2019, *arXiv:1905.11946*.

[38] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," 2015, *arXiv:1512.00567*.

QING SUN was born in Henan, China, in 1980. She received the M.E. degree in operations research and cybernetics from the Lanzhou University of Technology, Lanzhou, China, in 2008, and the Ph.D. degree in computational mathematics from Northwest Polytechnic University, Xi'an, China, in 2016.

She is currently a Lecturer with the Baoji University of Arts and Sciences. Her main research interests include computational mathematics and deep learning.

JING ZHAO was born in Shaanxi, China, in 1997. She received the B.S. degree in mathematics and applied mathematics from Weinan Normal University, Weinan, China, in 2022. She is currently pursuing the M.S. degree with the Baoji University of Arts and Sciences, Baoji, China.

Her main research interests include signal processing and target tracking.

CHUNLING XUE was born in Henan, China, in 1979. He received the M.E. degree in information and communication engineering from Northwest Polytechnic University, Xi'an, China, in 2010, and the Ph.D. degree in control theory and engineering from the Hi-Tech Institute of Xi'an, in 2023.

He is currently works at the Baoji University of Arts and Sciences. His main research interests include radar target detection and sea clutter suppression.

XUETING YANG was born in Shaanxi, China, in 1999. She received the B.S. degree in information and computing science from the Baoji University of Arts and Sciences, Baoji, China, in 2022, where she is currently pursuing the M.S. degree.

Her main research interests include signal processing and target tracking.

• • •